Development of low-stakes mathematics and literacy test þÿ s c o r e s d u r i n g l o w e r s e c o n d a r y s c h o o l A m u l t i l e pattern-centered analysis of student and classroom differences

Ketonen, Elina E.

2019-10

Ketonen , E E & Hotulainen , R 2019 , ' Development of low-stakes mathematics and literacy þÿ t e s t s c o r e s d u r i n g l o w e r s e c o n d a r y s c h o o l A m u l t i l e v e l p a t t e r n - c e n t e and classroom differences ' , Contemporary Educational Psychology , vol. 59 , 101793 . https://doi.org/10.1016/j.ced

http://hdl.handle.net/10138/304264
https://doi.org/10.1016/j.cedpsych.2019.101793

Contemporary Educational Psychology 59 (2019) 101793

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

Development of low-stakes mathematics and literacy test scores during lower secondary school – A multilevel pattern-centered analysis of student and classroom differences

Elina E. Ketonen*, Risto Hotulainen

Centre for Educational Assessment, Faculty of Educational Sciences, University of Helsinki, Finland

ARTICLE INFO

Keywords:
Classroom effect
Low-stakes assessment
Longitudinal study
Multilevel model
Pattern-centered analysis

ABSTRACT

The development of students' learning and test-taking behavior may derive from the social context and the group of peers they associate with daily for years. Consequently, it is assumed that students' academic achievements are to some degree affected by their classmates and the composition of the classroom. The present study provides evidence on how Finnish students (N = 5071) from different classrooms (N = 435) develop distinct patterns regarding their mathematics and literacy achievement during lower secondary school. We analysed longitudinal large-scale educational assessment data using a multilevel latent profile analysis (MLPA) to investigate the impact of classroom effect on students' achievement patterns, that is, on the development of students' low-stakes mathematics and literacy test scores from 7th to 9th grade. The results demonstrated the added value of modelling the multilevel structure inherent in educational assessment data: we identified four student achievement patterns that displayed different distributions across the school classes. More precisely, besides individual characteristics, the development of students' low-stakes mathematics and literacy test scores was associated with class-level factors and some of the classrooms seemed to have a stronger effect on students' test scores. These results suggest that classroom context is associated with students' achievement patterns, especially regarding the worst achieving students. The findings may reflect a combination of class placement practices as well as classroom and peer effect. Although the differences between Finnish schools have been one of the lowest in the OECD countries, the findings of the present study suggest that the classroom membership may create class level quality differences in both the preconditions and the development of learning.

1. Introduction

Educational situations are bound to a multilevel context: students are nested within their classrooms and schools. Consequently, the development of students' learning and achievement may also derive from the social context and the group of peers they associate with daily for years. Thus, it could be assumed that the development of a student's learning, test taking behaviour and consequently their academic achievements are to some degree affected by their classmates and the composition of the classroom. The objective of the present study was to investigate whether students from different classrooms develop distinct patterns regarding their mathematics and literacy achievement during lower secondary school.

It is well known that there is a range of ways that students are sorted into different tracks and/or classrooms in different educational systems, such as on the basis their ability, interest and their parents' deliberate choices, indirectly producing social reproduction (e.g., Eurydice, 2018; Reichelt, Collischon, & Eberl, 2019). Even in countries in which tracking options are not externally visible, some within-system related alternatives, including foreign language options, extra sports, arts or any other subject hours, steer students away from commonly shared learning experiences. Simultaneously such choices create unintended consequences related to students' social learning environments and teachers' expectations (Van den Broeck, Vansteenkiste, De Witte, Soenens, & Lens, 2010; Warburton, 2017) which in turn become measurable as class-level effects on achievement (Timmermans & Rubie-Davies, 2018).

Although innumerable studies can be found modelling the development of students' academic achievements, methods grouping students into different development patterns, achievement clusters or profiles based on their performance have generally ignored the classroom level in their models (e.g., Bowers & Sprott, 2012; Geiser,

* Corresponding author at: Faculty of Educational Sciences, P.O. Box 9, 00014 University of Helsinki, Finland.
E-mail address: elina.e.ketonen@helsinki.fi (E.E. Ketonen).

https://doi.org/10.1016/j.cedpsych.2019.101793

Available online 13 July 2019

Lehmann, & Eid, 2010; Hart et al., 2016; Hickendorff, Heiser, Van Putten, & Verhelst, 2009). Furthermore, earlier studies concentrated mainly on individual and class-level cross-sectional data points, omitting closer examination of how the abovementioned levels interact with each other over time. The present study deepens this knowledge by providing evidence on how lower secondary school classes differ from each other in terms of students' mathematics and literacy achievement based on their test results and their development from 7th to 9th grade.

1.1. Class placement and class composition

Students' placement in schools and classes according to their academic achievement is a common practice within school systems (Eurydice, 2018). In some Organisation for Economic Co-operation and Development (OECD) countries, students are sorted relatively early (before age 13) into separate tracks to follow separate curricula built for each track, independently leading to the acquisition of different competencies related to their further education and work (e.g., Austria, Belgium, the Czech republic, Germany, Hungary, the Netherlands, the Slovak Republic, Switzerland and Turkey; OECD, 2013) whereas in other OECD countries (e.g., Australia, Canada, Chile, New Zealand, Poland, the Nordic countries, the United Kingdom and the USA), all students belonging to the same age cohort follow more or less the same curriculum through their primary and lower-secondary years of schooling and even beyond (e.g., Eurydice, 2018; OECD, 2013). However, even within the latter system there are identifiable temporary or more permanent groupings of students leading to different classroom compositions based on students' abilities, interests, language choices and so forth.

Research on compositional effects describe the impact of groups on individual differences and development (Raudenbush & Bryk, 2002). Prior research in the educational field has mainly focused on the effects of ability composition on cognitive performance (e.g., Dicke et al., 2018; Harker & Thymms, 2004; Thrupp, Lauder, & Robinson, 2002). It has been suggested that a student will have better achievements if the average ability level of the class is higher, whereas lower achievement is expected among lower-performing peers (Peetsma, van der Veen, Koopman, & van Schooten, 2006; Rubie-Davies, 2009; Stäbler, Dumont, Becker, & Baumert, 2017). In addition, recent findings have shown that when the number of students with support needs within a class increased above a certain proportion it had a lowering effect on the overall class-level achievement (Hienonen, Lintuvuori, Jahnukainen, Hotulainen, & Vainikainen, 2018). In other words, the class composition may predict both the individual and overall performance level of the class.

The main argument for competence-based grouping comes from the teachers who often state that they are more able to respond the needs of homogenous study groups. However, the results of several studies have provided an ambiguous picture of the effect of homogeneous grouping on students' achievement (Forgasz, 2010; Slavin, 1990). Opposing arguments against homogeneous grouping are grounded partly on the Matthew Effect by stating that those who are not chosen for the higher track, in other words, among the better-achieving students, will suffer from the effect of low achieving peers, less demanding learning content and objectives, less competent teachers and lower teacher expectations (Brophy, 1983; Hanushek, Kain, & Rivkin, 2004; Hienonen et al., 2018; Jussim & Harber, 2005). This argument is often supported by the findings showing that the socio-economic background of students is closely tied to their educational performance, meaning that early streaming is likely to renew and even strengthen the existing differences between well-educated and less-well-educated families, thereby perpetuating inequality within society (Acacio-Claro, Doku, Koivusilta, & Rimpelä, 2018; Collins & Gan, 2013).

1.2. Classroom context and academic achievement

Sorting students either purposefully or unintentionally into different classrooms has inevitably led to at least two sets of consequences for students' learning. The first direct sorting effect based on ability or shared interest is related to teachers' expectations (Forgasz, 2010). Teachers tend to form expectations not only for the individual students but also for the groups of students, according to their beliefs about student learning and teaching (Rubien-Davies, 2010). The second consequence of sorting is related to peer effect. While class placement and study-related choices create differences between students' tangible reference groups, every selection made, or option given to a student and his or her parents also gradually builds the students' psychological awareness of these choices (Harter, 2012). In following, both effects will be described briefly in relation to the development of academic achievement.

1.2.1. Teacher expectations and classroom climate

There has been a long research tradition evidencing the teacher effect on student achievement (for a review, see Darling-Hammond, 2000). For instance, teachers' expectations have an effect on students' learning, academic engagement and achievements (Brophy, 1983; Rosenthal & Jacobson, 1968). According to Rubie-Davis (2009), teachers form their expectations of their class and students according to students' earlier performance or other available information and by their beliefs about teaching and learning. Teachers tend to expect and require more from those students who they perceive of having more learning potential producing the so-called self-fulling prophecy effect (Glock & Krolak-Schwerdt, 2013; McKown & Weinstein, 2008; Rosenthal & Jacobon, 1968; Rubie-Davis, 2009).

Recently, studies related to teacher expectation has focused more on class level effects while research has shown that teachers' expectation effect on individual students is rather small (Timmermans & Rubie-Davies, 2018). Teachers seem to have different expectations of achievement depending on the class composition. For instance, the general standard of learning and achievement within the class may be lower because of the presence of lower-performing students in the class (Goldenberg, 1992). Consequently, if teacher expectancy is low and the requirements of the whole class are less demanding, this may be manifested in students' actual performance. Interestingly, studies have shown that there is variation in how differently teachers' expectations work in different classes – in some classes, achievement gap between students increases whereas in other classes it decreases (McKown & Weinstein, 2008; Timmermans, Kuyper, & van der Werf, 2015, Timmermans & Rubie-Davies, 2018). In some cases, sorting students according to their abilities or interest may provide opportunities for the teachers to focus on a narrower range of learning needs causing the latter effect (Collins & Can, 2013).

Furthermore, teacher expectations have been found to be related to the creation of class climate (Rubie-Davis, 2009) which in turn is shown to have a significant impact on academic success (Thapa, Cohen, Guffey, & Higgins-D'Alessandro, 2013). For example, research has shown that when teachers work with students coming from low socioeconomic background they build a more structured environment, give students less autonomy and fewer engaging tasks (Glock & Krolak-Schwerdt, 2013; Van den Broeck et al., 2010). Class climate is a multi–faceted construct that reflects students' perceptions of their interactions with peers, teachers, and school administrators including shared school-related beliefs, values, and attitudes (Thapa et al., 2013). Previous studies have found that perceptions of both the peer and teacher-created climates influenced the quality of student motivation and effort in the classroom (Van den Broeck et al., 2010; Warburton, 2017; Wentzel, Muenks, McNeish, & Russell, 2017). It could be argued that especially in test situations, the teacher's expectancies and the values the classmates attach to the test influence students' engagement and achievement (Pintrich & DeGroot, 1990). These class-level effects can be intensified, particularly in low-stakes assessment situations in which the incentives and rewards might be absent, with the values playing a bigger role in performance (Attali, 2016; Cole, Bergin, & Whittaker, 2008; Van Barneveld, Pharand, Ruberto, & Haggarty, 2013).

1.2.2. Peer effect

Previous research indicates that in classroom settings, peers have an influence on student's learning, motivation, school adjustment, and achievement (Altermatt & Pomerantz, 2003; Kindermann, 2007). Classmates can provide information about what is valued and important in relation to particular tasks and activities and can therefore affect the motivation and engagement in the classroom as well. Consequently, peers may have a key role in shaping the social context and motivational climate at school and classes by providing cues and messages about what is valued in terms of competence and success (Harter, 2012). These cues and the structure of the classroom environment can influence whether a student is striving to perform well in most academic situations including low-stakes conditions. For instance, peer group membership has been related to school valuing (e.g., Ryan, 2001) and students' academic effort at school (e.g., Kindermann, 2007). Longitudinal evidence also supports the importance of peers in influencing adolescents' motivational outcomes (e.g., Makara & Madjar, 2015; Shin & Ryan, 2014), also referred to as the group contagion effect (e.g., Prinstein & Dodge, 2008) and enhancing internalization of academic behaviours (Harter, 2012).

In terms of achievement, previous studies have indicated that the student's achievement will be higher in a high achieving class than in a low achieving class (e.g., De Fraine, Van Damme, Van Landeghem, Opdenakker, & Onghena, 2003; Stäbler et al., 2017), implying a positive peer spillover effect (Willms, 1985). Lam's (2014) closer examination of effects of peer networks, such as friends, study mates, emotional supporters, and seatmates showed that all networks have an effect on school achievement (Lam, 2014). Positive network effects were stronger for friends and seatmates, especially, in major subjects like mathematics. While it has been stated that students may gain from similarly-able academic peers, unfortunately, among the low-achieving students, the effect is often stated as being a negative one (Tieso, 2003; see also Collins & Gan, 2013). It is hypothesized that whereas homogenous grouping eases a teacher's work and benefits high-achievers in the form of alike peers, the peer effect works in the opposing direction among low-achievers (cf. Collins & Gan, 2013). However, research findings are somewhat inconsistent regarding the positive peer spillover effect (see Dicke et al., 2018) and the overall compositional effect (for an overview, see Nash, 2003), and still inconclusive with respect to the effect of homogeneous grouping on students' achievement (Forgasz, 2010). We propose that a pattern-centered approach, simultaneously focusing on student characteristics as well as classroom differences, could extend the current understanding of peer group influence on student achievement.

1.3. The present study

1.3.1. Class placement during Finnish comprehensive school

In Finland, the differences in PISA results between schools were the second smallest among the OECD countries in 2015 assessment (OECD, 2016). Nevertheless, there are already considerable class-level differences at the 4th and 8th grade (Yang Hansen, Gustafsson, & Rosén, 2014) and such differences increase by the end of the nine-grades of comprehensive school (Hotulainen, 2016). In many school-effect studies, achievement differences within the schools have been found to be related to differences in class composition (e.g., Harker & Tymms, 2004), applying to Finnish context as well (Berisha & Seppänen, 2017; Hienonen et al., 2018). Regardless of the absence of both official ranking lists and a tracking system within the Finnish comprehensive school, parents may make deliberate choices when children proceed through the grade levels. In the Finnish system, some choices about study subjects and courses during the educational path (e.g., in music, foreign languages, sports etc.) may cause a cumulative effect on both school and class level clustering by the end of the Finnish comprehensive schooling.

Furthermore, the Finnish special education system (referred to as Learning and Schooling Support) has its own effect on the clustering. The system is based on the observed educational needs, not on medical diagnosis, and the main aim is to allocate support to the student rather than to bring the student to the support services (Jahnukainen & Itkonen, 2015). There are three levels in this support system: general (Tier 1), intensified (Tier 2) and special (Tier 3) support. The support methods and tools do not differentiate between tier levels, but the intensity of the support provided does (Finnish National Board of Education, 2014; Thuneberg et al., 2013; see also Jahnukainen & Itkonen, 2015). Tier 1 general support is targeted to the whole student population. Tier 2 intensified support is implemented when a student needs a longer period or more intensive support and 9.0 per cent of students receives this level of support. At the Tier 3 level, special support consists of the whole continuum of special education services and an administrative decision. An individual education plan is always required for students receiving Tier 3 support. A total of 7.5 per cent of all comprehensive school students received Tier 3 support and of those, half were placed in regular classes and the other half in special classes, or in special schools (OSF, 2017). In the Finnish school system, the culture of inclusion is very strong, and students of all kinds are usually placed in regular classrooms. Similarly, there are no special classes for "gifted" students.

In Finland, comprehensive school is a continuum including primary (grades 1–6) and lower secondary (grades 7–9) education, but in practice, many municipalities still have separate schools for them. Therefore, students are quite often allocated to new classrooms with new classmates at the beginning of the 7th grade. Although their core peer group remains the same most of the time, students have different subject-specific teachers during lower secondary school. In Finland, there is no official formal tracking and all the pupils are guaranteed equal educational opportunities during the whole comprehensive school, but in addition to the cumulative effects of subject choices, classes with a special emphasis provide implicit tracks within the system (Berisha & Seppänen, 2017; Koivuhovi, Vainikainen, Kalalahti, & Niemivirta, 2019). Furthermore, the increased diversity of school areas, free school choice policy (Kosunen, Bernelius, Seppänen, & Porkka, 2016) and increased class choice because of classes with a special emphasis (Koivuhovi et al., 2019) have begun to increase the achievement gap between classes and schools (Berisha & Seppänen, 2017; Hotulainen, 2016). In the present study, although we do not have exact information about the individual choices of single students and how they may have affected the students' class placement during their comprehensive school years, overall, the consideration of a broader social context in models of students' cognitive test responses seem to be reasonable.

1.3.2. Research questions

In the present study, we are interested in knowing how students' low-stakes mathematics and literacy test scores develop from the 7th to the 9th grade. As with earlier studies, we expect to find different achievement patterns (Hotulainen, 2016). However, to extend our knowledge, we analysed longitudinal large-scale educational assessment data using a multilevel pattern-centered approach to investigate the impact of classroom effect on students' mathematics and literacy achievement patterns. This approach was chosen because in addition to student achievement patterns, it also identifies patterns of classrooms and their association with classroom characteristics. A focus on the class level presents an important opportunity to detect those classrooms in which the effect of schooling and peers on student achievement and engagement might be stronger and therefore, may contribute to advancing our understanding of classroom differences. Although it is clear that the social context relates to achievement, little research exists that incorporates a pattern-centered analytical approach to examining classroom differences. Overall, this approach has the potential to identify effective strategies for targeting students and classrooms at risk for lowering achievement. Further, student achievement in aggregate may contribute to a cumulative, overarching achievement pattern at class-level that leads to a unique effect on student attitudes, behaviour and performance, manifested especially in low-stakes test situations. The following three research questions were addressed:

(1) What kinds of mathematics and literacy achievement patterns can be identified among lower secondary school students and how can they predict actual school performance at the end comprehensive school?

Based on prior research, we anticipated identifying subgroups of students with different levels of mathematics and literacy achievement (e.g., Bowers & Sprott, 2012; Hart et al., 2016). Earlier studies have shown that both increasing and decreasing patterns can be found (Bowers & Sprott, 2012; Hotulainen, 2016). Those students that were part of the increasing achievement patterns were expected to have better school leaving grades than those with decreasing test scores (Hotulainen, 2016).

(2) Are there differences between lower secondary school classrooms in the occurrence of distinct student achievement patterns (i.e., classroom patterns)?

Based on the indicated variance in different test achievements between Finnish classrooms (Thuneberg, Hautamäki, & Hotulainen, 2015; Yang Hansen et al., 2014), we expected to find different classroom patterns with a varying number of distinct student achievement types.

(3) Are there differences between classroom patterns in terms of how much of the variance in the development of students' achievement is explained by classroom membership?

Since achievement differences within the schools have been found to be related to differences in the class composition in many earlier studies (e.g., Berisha & Seppänen, 2017; Harker & Tymms, 2004; Hienonen et al., 2018), we suggest that the classroom patterns identified would play varying roles in the development of students' mathematics and literacy achievement.

2. Method

2.1. Participants

The present study used data from the Metropolitan Longitudinal study in Finland, conducted in lower-secondary schools in 14 municipalities in the Helsinki metropolitan area in southern Finland. The data were collected from all comprehensive schools in the area, with few exceptions. Five schools were omitted for the reason of refusal or technical issues. The students (and classes) were reached via the participating schools and all students in the target grade were asked to participate. At the student level, typical reasons for the nonresponse were students' absences from the class on the survey day or not having received the parents' consent statement. This type of missing data cannot be identified or managed in our dataset and but should be borne in mind when interpreting the results.

The study population in the present study completed identical achievement tests (in mathematics and literacy) in 2011 (7th graders, 12–13-year-olds) and 2014 (9th graders, 14–15-year-olds) and this data contained 5464 students (50.4% girls) from 504 classrooms and 117 schools in the Helsinki metropolitan area. In the analyses, only those students who stayed in the same class during their three years of lower secondary schooling were included (N = 5195). Furthermore, special education classes and those classrooms having less than five cases (124 students and 35 classrooms) were excluded from the analyses in order to ensure representative class-level results. The final group of participants consisted of 5071 students (50.7% girls) from 435 classrooms (between 5 and 22 cases per class) from 110 schools, all of whom stayed in the same class during their three years of lower secondary schooling. Within these data, the actual class size, which was extracted from the original ninth-grade student lists received from the Education Departments of the municipalities, ranged from 10 to 30 students per class (M = 19.9, SD = 3.1, Mode = 20).

In terms of students' socioeconomic status (SES), which was measured by mothers' educational level, 45.5% of participants reported that their mother had completed polytechnic- or university-level education, 34.6% high school and 19.9% basic education. Special education needs (SEN) was measured at the ninth grade by asking special education teachers to complete a questionnaire to provide information on whether students received support. With this criterion, 301 students were labelled as having SEN status (193 Tier 2 students and 108 Tier 3 students). The class composition measure was calculated according to the original class information (and size) by aggregating the received support to the class level as a mean percentage of the Tier 2 and Tier 3 students (Min = 0.0% , Max = 55.6%, M = 10.2%, SD = 9.9%, classes with SEN students n = 299). The percentage of SEN students and class size correlated negatively (r = –0.14, p = .006). Ethical approval for the data collection was obtained from the Ethical Committee of the National Institute of Health and Welfare. Permission to conduct data collections in schools was obtained from the Education Department of each municipality.

2.2. Materials

The assessment contained identical mathematics and literacy achievement tests at the beginning of lower secondary school in 7th grade and at the end of lower secondary education in 9th grade. Both data collections were conducted by teachers during an ordinary school day. In grade 7, the achievement tests were conducted using paper and pencil and at grade 9, tests were conducted electronically. The mathematics and literacy tests were modified versions of the official national curriculum-based mathematics and literacy tests covering the learning objectives of primary grades (Finnish National Board of Education, 2004). These tests measure the basic knowledge and skills students ought to acquire when entering lower secondary school. Although the tests were not high-stakes, it can be assumed that most students tried to do well, since the test situations were incorporated into normal schoolwork and tasks of this kind show their competence and are often related to their grading. For the analyses, the percentage of correctly-solved tasks was calculated according to the mathematics and literacy achievement tests.

The grade point average (GPA) was calculated as a mean for theoretical subjects, based on students' self-reports of their prior achievement in grade 7, and derived from the National Joint Application Register (comprehensive school diploma) in grade 9. The measurements for mathematics achievement correlated positively with both self-reported prior GPA (r = 0.54 and 0.49, $p < .001$) and register-based GPA at the end of lower-secondary school (r = 0.53 and 0.54, $p < .001$). Also, the measurements for literacy achievement correlated positively with both self-reported GPA (r = 0.52 and 0.48, $p < .001$) and register-based GPA (r = 0.53 and 0.55, $p < .001$).

2.3. Methodological approach

The pattern-centered approach is a technique that enables the study of person-specific configurations of mathematics and literacy achievement (also referred as a person-oriented approach; e.g., Bergman & Andersson, 2010). Pattern-centered methods represent a cluster analytical approach, one in which students with a similar profile in a set of variables can be classified as one type (e.g., Vermunt & Magidson, 2002). Amongst the pattern-centered methods, latent profile analysis (LPA) has the advantage that it represents a model-based approach which allows evaluation of the model fit and the comparison of different models with distinct numbers of profiles (Vermunt & Magidson, 2002). In educational research, LPA has previously been applied to identify patterns of students' cognitive strategy use (Fagginger Auer, Hickendorff, Van Putten, Béguin, & Heiser, 2016; Geiser et al., 2010; Hickendorff et al., 2009), study engagement profiles (Ketonen et al., 2016) and homework learning types (Flunger et al., 2015).

However, LPA assumes that observations are independent of each other and therefore less suitable for use with hierarchical datasets in which individuals are nested within higher-level groups (e.g., classrooms). Furthermore, ignoring the presence of a nested data structure may lead to inaccurate classification of individuals in addition to violating the independency of observations assumption (Asparouhov & Muthén, 2008). Because of the inherent nested structure of educational datasets, the LPA statistical technique has been extended to adopt an additional hierarchical level (Henry & Muthén, 2010). Multilevel latent profile analysis (MLPA) considers the different levels of data and thus enables the investigation of research questions that are beyond the scope of single-level LPA. For instance, qualitatively-different classrooms can be identified and further investigated, along with student-level patterns.

At the lower student level, different student profiles can be detected (e.g., patterns of student achievement). In addition to the nesting of variables within individuals, the nesting of individuals in higher-level units is included in the model. The student-level profiles may vary across classrooms and this variation in the higher level can be modelled with class-level latent profiles (i.e., patterns of classrooms). In non-parametric MLPA, the indicators for the class-level profiles are the student-level profiles themselves. In this approach, the random means from the Level 1 latent profile solution are used as indicators of a second latent profile model at Level 2. The different Level 2 latent profiles have different distributions of the random means; that is, the log-odds of membership in a particular Level 1 latent profile. If class membership affects student-level profile membership, the distribution of student-level profiles among classrooms may significantly vary. Thus, MLPA may produce a higher order set of profiles that represents the distinct distributions of student-level profiles within classrooms (see Asparouhov & Muthen, 2008; Henry & Muthén, 2010). For instance, if a latent profile at the class level is characterized by a high number and high-class probabilities of a certain student-level profile (e.g., low achievers), this can count as an indication of a strong class effect on the formation of this specific achievement pattern.

Recently, the validity of the statistical models used in support of peer effects are challenged and stronger controls of measurement error and pre-existing differences are recommended to avoid spurious class effects (Dicke et al., 2018; Harker & Tymms, 2004). Although MLPA cannot be directly compared to models applying variable-oriented methods, it resembles a doubly latent model of contextual effects often applied in a SEM framework (e.g., Marsh et al., 2009). First, both models are latent since they use multiple indicators that are designed to reflect a hypothetical unobserved construct that is assumed to be responsible for the covariance among the indicators (i.e., latent profiles in the case of MLPA), thus allowing for the control of measurement error. Second, these models simultaneously control for sampling error by considering the clustered nature of the data. Furthermore, besides modeling the measurement error in the observed indicators and correctly modeling the nested structure of multilevel data, MLPA permits the simultaneous assessment of both individual (Level 1) and contextual (Level 2) predictors (Henry & Muthén, 2010). This feature allows for the possibility that individuals with the same Level 1 covariate values might differ in their probability to belong to a certain latent profile due to contextual or environmental differences in their community.

Within the variable-centered, doubly-latent modeling framework (e.g., MSEM), the group-level effects can distinguish between climate and contextual effects (e.g., Marsh et al., 2012). In the climate effect, class-average ratings are based on students' responses when students are asked to rate directly some characteristics of the higher-level unit to which they are exposed (Marsh et al., 2012). In the contextual effect, the referent is the individual student rather than the class, and class averages are used to describe classroom composition (i.e., context). Particularly in contextual studies, the same level 1 variable can be used to form constructs on both levels, but their interpretations may differ (see e.g., Marsh et al., 2009). In the present study, the latent construct at class-level is in a sense analogous to a contextual effect, since the make-up of the classroom latent profiles were based on individual-level achievement. To sum up, MLPA has the potential to provide knowledge on whether class-level variability exists in within-person achievement patterns, how classrooms can be clustered according to the proportions of certain student profiles and the extent to which achievement can be explained by (pre-existing) individual characteristics versus contextual factors (e.g., class composition). Given the emphasis in understanding possible inequality in education, this relatively new method allowing for the assessment of contextual-level predictors of individual typologies of behavior can have high relevance for practical implications and educational policy. This study makes a significant contribution to the literature by presenting an empirical example how the technique can be applied in the educational field.

2.4. Statistical analyses

In the present study, we investigated whether classroom membership plays a role in the development of students' distinct achievement patterns. For this purpose, we estimated MLPA model considering student and class-level covariates. The chosen statistical method allowed us to use several student achievement indicators simultaneously, reveal the dependence of the student types on their classrooms and to explore the associations with student and classroom characteristics. At the student level, we assessed both mathematics and literacy achievement at two time points (at the beginning and end of lower secondary school) and considered various students characteristics (i.e., gender, socio-economic status, prior achievement and special education needs (SEN) status). At the class level, the classroom membership during the lower secondary school as well as the class composition (i.e., class size and proportion of students with special education needs) were considered.

The MLPA was conducted with version 5.1 of the Latent GOLD program (Vermunt & Magidson, 2013). First, we ran a series of LPAs to estimate latent profiles of students who are characterized by a particular pattern of mathematics and literacy achievement. The model was then extended with a multilevel component by adding a latent class-level variable, on which students' probability of being in each latent student profile is dependent. All four mathematics and literacy test score variables were entered as observed response variables at Level 1 and a classroom identifier variable as the grouping variable for the multilevel effect (see Fig. 1 for graphical representation of the measurement model). The optimal number of student-level profiles was selected according to information criteria that consider model fit and complexity simultaneously: Akaike and Bayesian Information Criteria (AIC, AIC3 and BIC). At the class level, the best fitting model and the number of classroom profiles was selected by using a group-based BIC, as suggested by Lukočienė and Vermunt (2010) when having a multi-level design.

Next, the student-level covariates (gender, SES, prior achievement, SEN status), class-level covariates (class size and percentage of SEN students) and the GPA as the outcome variable at the end of lower-secondary education were added to the model (see Fig. 2 for the structural model). To explore the association of covariates with the classification into certain student- and class-level profiles, the three-step approach was chosen (Bakk, Tekle, & Vermunt, 2013; Vermunt, 2010). First, the latent profiles were estimated at both levels by taking into account solely the indicator variables of mathematics and literacy achievement. Second, individuals and classrooms were classified into their most likely latent profile membership (i.e., student and classroom types), relying on the most likely latent profile membership while considering the classification error probabilities as weights (modal

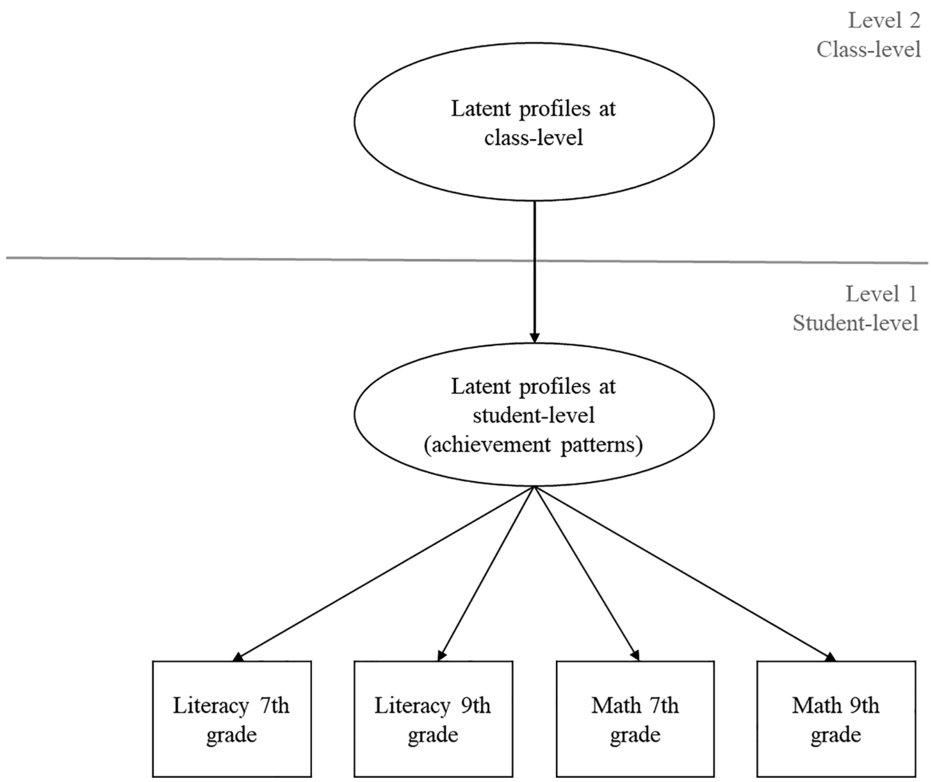

Fig. 1. Measurement model.

assignment) due to misclassification (e.g., Bakk et al., 2013). Third, external variables were considered to be predictors of the latent profile variables in a multinomial logistic regression model in which the classification errors in the outcome variables are considered in the estimation to prevent bias (see Bakk et al., 2013). The covariates were considered jointly in the models, in order to explore which student (or classroom) characteristic was most strongly associated with the achievement profiles while controlling for the associations with the other covariates. The final model was investigated in more detail by evaluating the statistical significance of each covariate and outcome

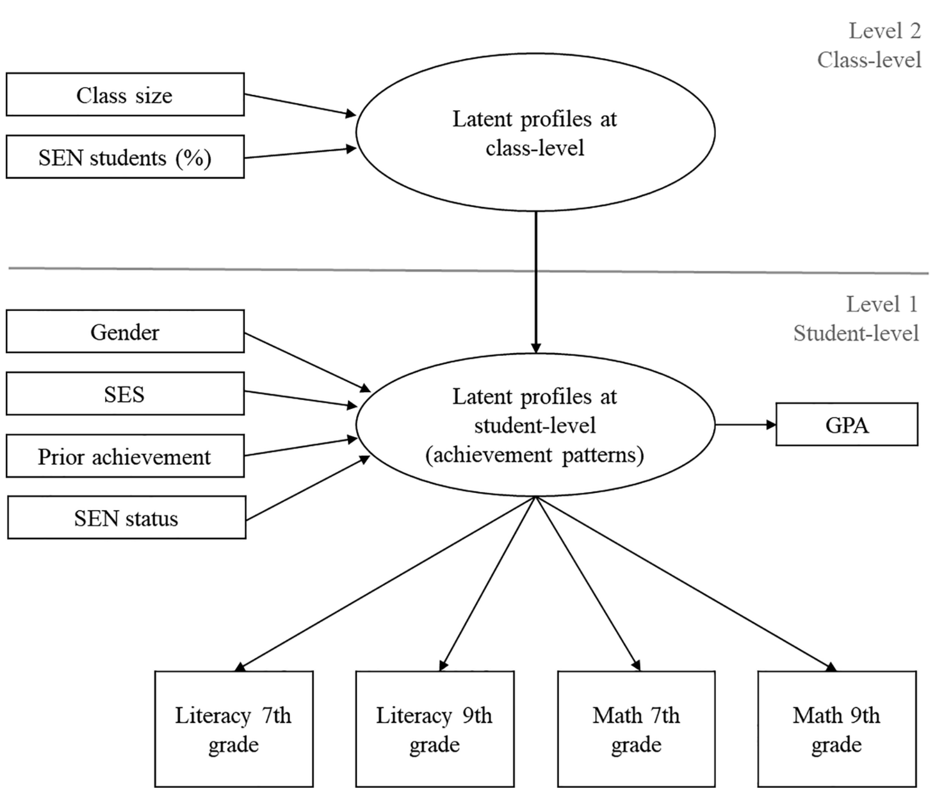

Fig. 2. Structural model.

variable with a Wald test.

To answer the third research question, the group-specific intraclass correlation coefficients (ICCs) in mathematics and literacy achievement change scores were calculated for each classroom profile. For this purpose, each classroom was accorded the most likely class-level profile based on the posterior probabilities of the MLPA. Multigroup ICCs and all preliminary results were calculated using Mplus version 8.1 (Muthén & Muthén, 1998-2018).

3. Results

Overall, the test scores in mathematics from 7th grade (M = 56.0, SD = 21.6) to 9th grade (M = 54.2, SD = 24.5) dropped slightly (t(5070) = 5.77, p < .001), whereas in literacy, the test scores increased (t(5070) = –14.38, p < .001) from 7th (M = 63.8, SD = 19.6) to 9th grade (M = 68.3, SD = 24.0). In the preliminary analyses, we also explored the intraclass correlation coefficients (ICCs) of 7th and 9th grade mathematics and literacy test scores at both the class-level and school-level. Since the ICCs at the school level were only 5% in both grades 7 and 9 (expressing the relative percentage of how much of the total variation in test scores is explained by the differences between schools), we decided to continue with a two-level model (students nested within classrooms). When considering only the student- and class-level variation in the 7th grade, 17.0% of the variation in test scores was explained by the differences between classrooms in mathematics, and 13.1% in literacy. In the 9th grade, the equivalent figures were 21.9% in mathematics and 16.4% in literacy. Expectedly, most of the variance in both mathematics and literacy scores was explained by individual factors, but the classroom effect seemed to be increasing from 7th to 9th grade, giving support to the hypothesis that the development of students' achievement and test taking behavior might be impacted by classroom effect. This was also supported by the ICCs in achievement change scores both in mathematics (16.1%) and literacy (9.9%) (indicating the relative percentage of how much of the total variation in the development of test scores is explained by the differences between classrooms).

Next, we ran a series of LPAs and MLPAs. Models with latent structures with up to seven latent student profiles and seven latent classroom profiles were fitted. Fig. 3 presents the elbow plot of the information criteria (AICs and BIC) for the different profile solutions of the single-level LPA. The point after which the slope flattens out indicates the optimal number of profiles in the data (Mäkikangas et al., 2018). As illustrated, between the one- and four-profile solutions, all the information criteria decreased, but then remained at the same level

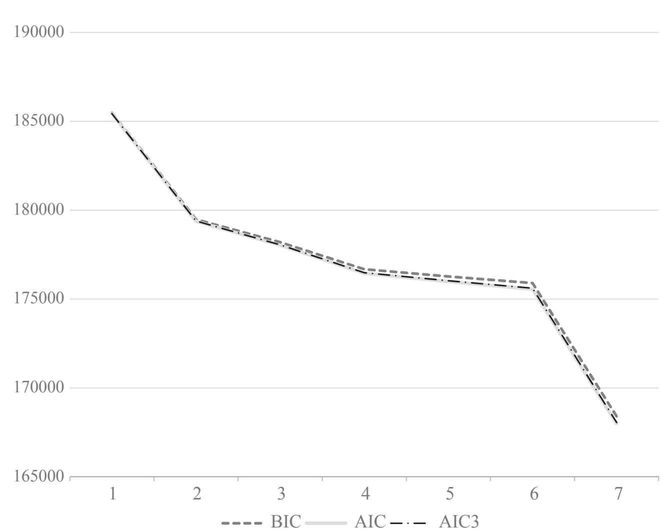

Fig. 3. Elbow plot of information criteria for different Level 1 solutions.

up to the six-profile solution. After the six-profile solution, the information criteria again decreased compared with previous solutions. Of the four- and seven-profile solutions, the model with four student profiles had a clearer interpretation and contained profiles with big enough memberships (i.e., > 5% of the cases). The four student-level profiles solution is shown in Fig. 4 (for profile-specific achievement test score means, see Table 2). The following profiles based on repeated achievement measurements were found: P1: *low and decreasing* (24.3%), P2: *below average and increasing* (45.1%), P3: *above average and increasing* (22.0%) and P4: *high literacy, above average mathematics and increasing* (8.6%). All three of the increasing profiles (P2, P3 and P4) had similarly modest development in their mathematics test scores based on the 7th grade starting values, with P3 and P4 having similar, higher results. In literacy, the increase in test scores was again related to the initial level of achievement varying from low (P2) to moderate (P3) and high (P4). However, literacy scores increased more than the mathematics scores among all three profiles with profile P4 gaining the most compared to other two increasing profiles. In the decreasing profile P1, the decline in test scores was notable in both subjects.

In the next step, the MLPA results showed that adding a multilevel structure greatly improved model fit (lower AICs and BICs), indicating a considerable within-class dependency of observations. While the AICs support a more complex model (six latent class-level profiles), the group-based BIC identified the model with four latent class-level profiles (with the already-defined four latent student-level profiles) as optimal (see Table 1). The proportional distributions of the student profiles varied significantly between classrooms, suggesting that the typicality of the achievement types varied from one classroom to the next (see Fig. 5, for detailed student profile probabilities in each latent classroom profile, see Table 3). The most common classroom profile C1 (N = 241) consisted of students from all four achievement profiles but had relatively higher proportions of *below average and increasing* (52.5%) and *low and decreasing* profiles (28.0%). The second classroom profile C2 (N = 121) was characterized by substantial probabilities for all achievement profiles except the *low and decreasing* achievers (only 8.7%). In the third classroom profile C3 (N = 53) the great majority of students were characterized by the profile *low and decreasing* achievers (65.1%) combined with a smaller proportion of students from the *below average and increasing* profile (25.2%). Finally, the fourth least-occurring classroom profile C4 (N = 20) consisted mainly of students from the *above average and increasing* profile (64.7%) and *high literacy, above average mathematics and increasing* profile (26.1%). Accordingly, we labelled the four divergent class-level patterns as classrooms having mainly: *increasing and decreasing low achievers* (C1, 55.4%), *increasing achievers* (C2, 27.8%), *decreasing low achievers* (C3, 12.2%), and *increasing high achievers* (C4, 4.6%).

In the third step, the student and class-level covariates and dependent variable of GPA were added to the model following the three-step approach. Gender, SES, prior achievement and SEN status were added to Level 1 as predictors of students' probability of being in a particular latent achievement profile and 9th grade GPA was regressed on the latent profile membership. At Level 2, class size and proportion of SEN students were added as predictors of the latent classroom profiles (see Fig. 2). The statistical significance and the profile specific probabilities or profile specific means of the covariates and dependent variable are illustrated in Tables 4 and 5. The Wald statistics indicate that student profiles differ statistically significantly in terms of gender (p < .001), prior achievement (p < .001) and SEN status (p < .001), but not regarding SES (p = .240; see Table 4). The proportion of boys in the *low and decreasing profile* was higher than in any other profile (63.0%). Furthermore, compared to the *above average and increasing profile* (53.0% girls), the proportion of girls in the *high literacy, above average mathematics and increasing profile* was even higher (62.7%). In terms of prior achievement, all the profiles differed from each other: students in the *high literacy, above average mathematics and increasing profile* having the highest achievement in terms of self-reported GPA (9.1), the *above*

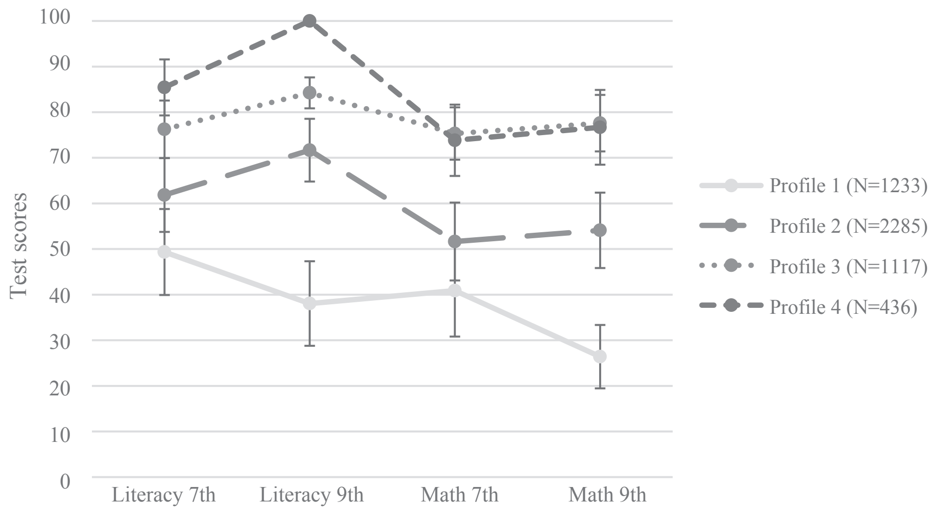

Fig. 4. Level 1 four-profile solution based on raw scores (with standard deviation of each mean).

average and increasing profile (8.9) and the *below average and increasing profile* (8.2) in between, and the *low and decreasing profile* reporting the lowest achievement (7.8). In terms of SEN status, the *low and decreasing profile* clearly differed from all the other profiles by having more students receiving special education support (18.8%; which is 73.3% of all SEN students in the sample). A smaller but significant difference was also found between the *below average and increasing profile* and the *above average and increasing profile*, the latter having relatively fewer SEN students (0.6%) than the former (3.5%). Finally, after controlling for all the covariates, the mean scores in 9th grade GPA still showed a clear pattern of differential associations with the distinct achievement profiles ($p < .001$). Students in the *high literacy, above average mathematics and increasing profile* had the highest GPA (9.2), the *above average and increasing profile* (9.0) and the *below average and increasing profile* (8.1) in between, and the *low and decreasing profile* had the lowest GPA when completing lower secondary school (7.3).

Concerning the classroom profiles, statistically significant differences were found both in terms of class size ($p = .004$) and especially in class composition ($p < .001$). The results (see Table 5) indicated that especially in classrooms having mainly *decreasing low achievers* (C3), the class sizes were smaller compared to all other profiles ($M = 19.1$) and the proportions of SEN students within the class, the highest ($M = 12.4\%$). Compared to this classroom type, classrooms having *increasing and decreasing low achievers* (C1) did not differ in terms of the proportion of SEN students ($M = 10.4\%$) but were slightly bigger in size ($M = 20.3$). On the other hand, within classrooms consisting mainly of *increasing high achievers* (C4) the proportion of SEN students was clearly the smallest ($M = 2.9\%$) and second smallest in *increasing achievers* –classroom profile ($M = 7.4\%$), class sizes being the largest in these two class-level profiles ($M = 21.8$ and 20.8).

Finally, to examine whether the classroom effect in the development of achievement was stronger in some of the class-level profiles, the group-specific ICCs in mathematics and literacy achievement change scores were calculated for each classroom profiles. The results (see Table 6) indicated that especially in those classrooms having mainly *decreasing low achievers* (C3) and *increasing and decreasing low achievers* (C1), the ICCs of change scores both in literacy (C3: 5.4%, C1: 6.9%) and especially in mathematics (C3: 14.1%, C1: 13.4%) were rather

Table 1
Fit statistics for the multilevel latent profile models.

Latent profiles					BIC	
Class-level	Student-level	Log-likelihood	AIC	AIC3	Individual-based	Group-based
1 (no multilevel effect)	2	−89670.9411	179375.8819	179392.8819	179487.0708	
	3	−88989.2175	178030.4351	178056.4351	178200.4885	
	4	−88190.5534	176451.1068	176486.1068	176680.0249	
2	2	−89475.7302	178989.4604	179008.4604	179113.7303	179067.2383
	3	−88751.6274	177561.2547	177590.2547	177750.9298	177679.9683
	4	−87966.8044	176011.6087	176050.6087	176266.6891	176171.2581
3	2	−89437.3987	178916.7974	178937.7974	179054.1483	179002.7623
	3	−88677.9172	177419.8344	177451.8344	177629.1311	177550.8287
	4	−87902.3504	175890.7008	175933.7008	176171.9431	176066.7243
4	2	−89431.6253	178909.2506	178932.2506	179059.6825	179003.4027
	3	−88662.7629	177395.5258	177430.5258	177624.4441	177538.8007
	4	−87884.8812	175863.7624	175910.7624	**176171.1668**	**176056.1602**
5	2	−89431.4837	178912.9674	178937.9674	179076.4804	179015.3066
	3	−88658.0032	177392.0064	177430.0064	177640.5461	177547.5621
	4	−87876.6492	175855.2984	175906.2984	176188.8649	176064.0704
6	2	−89431.4784	178916.9568	178943.9568	179093.5508	179027.4832
	3	−88652.2791	177386.5583	177427.5583	177654.7195	177554.3946
	4	−87869.9884	**175849.9769**	**175904.9769**	176209.7054	176075.1232
7	2	−89431.4513	178920.9026	178949.9026	179110.5776	179039.6161
	3	−88652.0688	177392.1375	177436.1375	177679.9204	177572.2546
	4	−87868.9561	175855.9119	175914.9119	176241.8025	176097.4325

Note. The lowest BICs and AICs are in boldface.

Table 2
Means of literacy and mathematics achievement test scores on 7th and 9th grade for each latent student-level profiles.

Achievement test	Student profile: Low and decreasing (*N* = 1233)	Below average and increasing (*N* = 2285)	Above average and increasing (*N* = 1117)	High literacy, above average mathematics and increasing (*N* = 436)
Literacy 7th grade	49.35	61.84	76.26	85.43
Literacy 9th grade	38.05	71.67	84.24	100.00
Math 7th grade	40.89	51.65	75.33_a	73.84_a
Math 9th grade	26.40	54.10	77.60_a	76.69_a

Note. Means within a row sharing the same subscripts are not significantly different at the $p < .05$ level.

high. Also in those classrooms including quite heterogeneous composition of mostly *increasing achievers* (C2), the development of mathematics test scores was related to classroom membership (ICC: 9.3%), whereas in classrooms consisting mainly of *increasing high achievers* (C4), the development of both mathematics (ICC: 2.3%) and literacy (ICC: 1.2%) achievement was almost totally explained by individual factors rather than differences between classrooms.

4. Discussion

The present study demonstrated the added value of modeling the multilevel structure inherent in large-scale educational assessment data: we identified four student achievement patterns that displayed different distributions across the school classes. These results suggest a meaningful association between classroom membership and student achievement patterns. More precisely, this study showed that the development of academic achievement in mathematics and literacy and related test taking behavior is associated with class level factors and in some of the classrooms there seem to be a stronger effect on the development of achievement. By applying a pattern-centered approach, it was discovered that the context and classmates seem to play a bigger role in those classrooms having more low achieving students, indicating a potential risk for increasing achievement gap between lower secondary school classrooms.

The results regarding student achievement profiles were largely in line with previous research (e.g., Bowers & Sprott, 2012; Hotulainen, 2016): most students had slightly increasing results (profiles P2, P3 and P4) but there were also students with notably declining results (P1). Especially worrying is the latter group of students, representing almost one-quarter of the sample. However, the decline in test scores may also indicate a lack of motivation in low-stakes conditions besides actual decline in performance. That is, if students do not perceive the importance or usefulness of an exam, their effort suffers and so does their test performance (Attali, 2016; Cole et al., 2008; Van Barneveld et al. 2013). This may also explain why the overall test scores in mathematics dropped slightly across the whole sample during lower secondary school. Another reason could relate to the fact that the test itself covers the curriculum objectives of mathematics and literacy in primary education and is not repeated as such during the lower secondary school. This may also cause a decrease of motivation among some students, especially, when they notice that they are not able solve the tasks successfully. Nevertheless, it is noteworthy that the *low and decreasing achievers* were clearly more prevalent in some classrooms than others. Consequently, it could be argued that in test situation, the teacher's behavior and expectancy and the influence of peers may have an additional effect on students' behavior (Altermatt & Pomerantz, 2003;

Table 3
Student profile probabilities in each of the four latent class-level profiles.

Classroom profile	Student profile probability: Low and decreasing	Below average and increasing	Above average and increasing	High literacy, above average mathematics and increasing
C1 (*N* = 241)	**28.03**	**52.54**	14.14	5.29
C2 (*N* = 121)	8.74	**46.10**	**32.40**	12.76
C3 (*N* = 53)	**65.14**	25.23	8.09	1.53
C4 (*N* = 20)	0.05	9.11	**64.72**	**26.12**

Note. The highest student profile probabilities within a latent classroom profile in boldface.

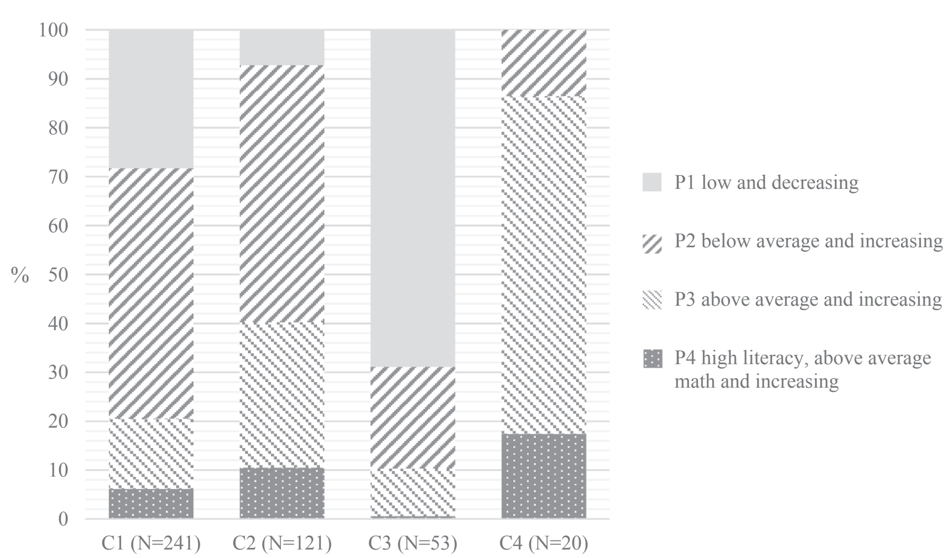

Fig. 5. Level 2 profiles based on the relative frequency of the Level 1 profiles.

Table 4
Profile specific probabilities or means of the covariate and dependent variables on student-level.

Variable	Student profile: Low and decreasing	Below average and increasing	Above average and increasing	High literacy, above average math and increasing	Wald-test	P
Gender					53.27	< .001
boys	63.02	44.95_{ab}	46.96_{a}	37.31_{b}		
girls	36.98	55.05_{ab}	53.04_{a}	62.69_{b}		
SES (mothers' education level)					11.48	.24
basic	26.72	20.98	11.22	7.37		
high school	35.72	34.75	31.44	26.39		
lower academic	18.31	23.21	24.71	26.78		
higher academic	14.77	17.92	30.91	38.06		
SEN status					44.18	< .001
no	75.82	88.75_{a}	90.99_{b}	88.85_{ab}		
yes	17.51	3.23_{a}	0.55_{b}	0.58_{ab}		
Prior achievement	7.6	8.2	8.9	9.1	445.69	< .001
GPA	7.3	8.1	9.0	9.2	320.53	< .001

Note. Profiles sharing the same subscripts within a row are not significantly different at the $p < .05$ level.

Table 5
Profile specific means of the covariate variables on class-level.

Variable	Classroom profile: C1	C2	C3	C4	Wald-test	p
Class size	20.3_{a}	20.8_{ab}	19.1	21.8_{b}	13.61	.004
SEN students (%)	10.49_{a}	7.39	12.35_{a}	2.87	28.98	< .001

Note. Profiles sharing the same subscripts within a row are not significantly different at the $p < .05$ level.

Table 6
Profile specific intraclass correlations (ICCs) of mathematics and literacy chance scores.

	Classroom profile ICC: C1	C2	C3	C4
Literacy development	**0.069**	0.038	**0.054**	0.012
Math development	**0.134**	**0.093**	**0.141**	0.023

Note. The highest ICCs of latent classroom profiles in boldface.

Goldenberg, 1992; Kindermann, 2007; Lam, 2014; McKown & Weinstein, 2008; Rubie-Davis, 2009; Ryan, 2001; Wentzel et al., 2017).

Even if low-stakes, the test scores correlated in an expected way with both self-reported prior achievement as well as register-based end of school GPA. Furthermore, the present study demonstrated the role of student characteristics in the test-scores-based achievement profiles by considering gender, SES, special education needs and prior achievement at the student level. Consistent with previous literature (e.g., Gibb, Fergusson, & Horwood, 2008), boys were more likely to belong to the *low and decreasing* achievement profile, whereas girls more probably represented especially the *high literacy* achievement profile. Furthermore, the *low and decreasing profile* clearly differed from all the other profiles by having more students receiving special education support. After controlling for all student characteristics (including prior achievement), the *high literacy* achievement profile succeeded the best at the end of lower secondary school in terms of GPA, whereas *low and decreasing* achievement profile performed least well. Significant differences were also found between the *below average and increasing profile* and the *above average and increasing profile*, the latter having relatively fewer SEN students and receiving better school leaving grades than the former.

The second finding of the present study was how the class-level clustering was related to the probability that a student will be a member of a particular achievement profile. Modeling the multilevel structure for the classroom level clearly improved the model fit and four latent classroom profiles were found with notable differences in the proportions of certain student profiles. The range was largest for the *low and decreasing student profile* and smallest for the *below average and increasing profile* (most evenly spread in every classroom type). Therefore, classroom context seems to be associated with students' achievement patterns, especially regarding the best and worst achieving students. However, prior achievement and the level of test scores in 7th grade also seemed to relate to the student profiles (and development of achievement) to some extent, since the order of the achievement levels between student profiles remained the same. Together these findings may indicate implicit tracking at the beginning of Finnish lower secondary school (see also Berisha & Seppänen, 2017; Kosunen et al., 2016), which in turn may result in cognitively less or more demanding classroom settings becoming visible in 9th grade test scores (Cole et al., 2008). Thus, the findings may reflect a combination of class placement practices as well as classroom effect.

However, the starting point (7th grade test scores) did not explain all the variation in classroom profiles and temporal differences were found. The final, very important finding was that in those classrooms having mainly *increasing and decreasing low achievers* (C1 and C3), the change in literacy and especially in mathematics achievement was more strongly-related to the class-level factors than in other classroom types. This indicates that the role of classroom and peers might be bigger in terms of development of achievement and test taking behaviour especially in those classrooms having proportionally more SEN-students and low achievers at the beginning of lower secondary school. Thus, are we hurting low-performing students by sorting them into lower ability groups? Are those students showing decreasing test scores hampered by their similarly-able academic peers and thus more vulnerable to class composition effect (see also Collins & Gan, 2013; Tieso, 2003)? In addition to possible peer effect, the increasing achievement gap between classrooms may be due to variations in how teachers' expectations work across different classes (McKown & Weinstein, 2008; Timmermans et al., 2015; Timmermans & Rubie-Davies, 2018). However, in these classrooms there were also low-achieving students showing an increasing pattern in test scores. With these students the finding could also be interpreted as evidence of the positive influence of schooling and peers: the classroom context supports the positive development of the low performers (Collins & Gan, 2013).

To sum up, it seems that the deliberate or unintentional class placement practices are creating further educational inequality in terms of heterogeneous compositional effects: in those classrooms having mostly high achievers, almost all the variance in test score development is explained by individual factors rather than the classroom context. On the other hand, in those classes having mainly low performers or more heterogeneity in terms of achievement, the contextual effect is clearly stronger and increasing towards the end of comprehensive school but could be inferred as a positive or negative influence of schooling and peers, depending on the student level pattern. Considering these kinds of differential effects resulted in a more comprehensive understanding of students' achievement and test taking behavior during lower secondary school, revealing the role of contextual factors on academic achievement in addition to the student characteristics. Besides indicating feasible achievement gaps between lower secondary school classrooms, consideration of low-stakes test scores may be an important window into students' engagement and internalization of academic behaviours. Future research should explore how sensitive low-stakes tests are to motivational bias (Van Barneveld et al., 2013).

4.1. Limitations

Contrary to many high-stakes tests, such as PISA or final exams, mathematics and literacy achievement was measured in the present study with tests that might have been perceived as low-stakes tests by students and might therefore present a mixture of both achievement and motivational factors. Although the measurements correlated relatively strongly with both prior achievement and end of compulsory school GPA, it is still important to test the generalizability of our results using varying measures of achievement (e.g., both high and low-stakes tests). Furthermore, since we didn't use a modeling approach that could explicitly capture longitudinal changes, the developmental differences may have been understated in the MLPA model. Some of the development of the test scores in literacy and mathematics may be related to the change to from the paper/pencil test format used in the 7th grade, to the electronic test used in grade 9, which according to PISA studies, has been shown to lead to lower scores (Jerrim, 2016). However, according to previous assessments, differences between paper and online versions have been relatively small and comparable in Finnish samples (Hautamäki, Kupiainen, Marjanen, Vainikainen, & Hotulainen, 2013). Overall, since the tests were not particularly planned for growth assessment, and the interest was in detecting student patterns rather than trajectories, the reader needs be cautious about interpreting students' learning and its advancement on the basis of these results.

It should be noted that the classroom profiles identified do not reflect the effects of the teachers in our sample, since in Finland students have different subject-specific teachers during lower secondary school, although their core peer group remains the same most of the time. Furthermore, since class composition tends to vary slightly by subject area in lower secondary education, the time students spend in their classes can vary for each student during a school day. Consequently, the results could be different to some extent in countries having different educational systems. Furthermore, we only included those students who were present at school on data collection days and students chronically absent may be missing from the analysis. As another limitation, data about possible field of specialization within classrooms (or schools) is unfortunately not available in the present study as they might have shed more light on the class-level findings. For future research and practical implications, it would be important to collect detailed class-level data that could explain the mechanisms of how the classroom effects are mediated by various classroom characteristics and practices (e.g., teacher expectations). Importantly, the results of this study do not allow the causal direction of the association between student achievement patterns and classroom context to be concluded. It may be that certain classroom contexts lead to declining performance, but the initial levels of students' achievement may also play a role in class placement practices as well as in the development of the classroom learning climate. Future research should identify the directional associations between achievement development and classroom context, as well as detect how different classroom characteristics influence students' progress.

4.2. Implications

The results indicate the existence of significant differences between classrooms for both mathematics and literacy test scores. It appeared that even more of the variance in 9th grade mathematics scores was attributable to differences between classes than at 7th grade - nearly 22% compared with 17%. This was in line with the findings for literacy achievement (nearly 16.5% compared to 13%) but the trend was more marked in mathematics. Some information about reasonable values of intraclass correlation coefficients (ICCs) comes from cluster-randomized trials that have been conducted. For example, Hedges and Hedberg (2007) presented a compendium of intraclass correlations for academic achievement. While there is conventional wisdom suggesting that 'effective enough' values of these ICCs are typically between 0.05 and 0.15, they suggest that values between 0.15 and 0.25 are more plausible in terms of significant statistical power for achievement outcomes (Hedges & Hedberg, 2007). Thus, even within this very restrictive frame of reference, the ICCs of the present study revealed a substantive and significant amount of class-to-class variation in mathematics and literacy achievement.

While the overall change in students' test scores was clearly related to classroom membership, from a student's (or parents') perspective, not all the classrooms are the same in terms of development of test scores. In those quite homogeneous classrooms consisting mainly of high achievers, only 1–2% of the development of mathematics and literacy achievement scores was explained by the context (classroom membership). However, in classrooms including more low performers and heterogeneity, the contextual effect (both positive and negative) was clearly stronger especially in mathematics, explaining over 13% of the change in test scores regarding two of the classroom types. Our main point, however, is not the effect sizes, but that the magnitude of class effect might be different for both students and classrooms of a different type, which is most important result from a policy perspective.

Although it is impossible to fully consider the pre-existing differences between classrooms (Harker & Tymms, 2004), the reported intraclass correlation coefficients of different classroom profiles focused on the change scores rather than the initial achievement levels, thus, revealing the possible differences in the magnitude of classroom effect regardless of the inevitable pre-existing differences in achievement due to class placement practices. These results have significant implications for the schools and parents. For instance, it is misleading to use 'raw' rather than 'value-added' test results as measures of classroom (or school) effectiveness in promoting pupils' academic achievement. Our findings indicated that within those classrooms with large proportion of high achieving students, the context plays a clearly smaller role in the development of test scores. On the other hand, for low achievers the classroom context and classmates may more likely either hinder or facilitate academic achievement, implying a differential effect (see also Sammons, Nuttall, & Cuttance, 1993). In these classrooms, there were also proportionally more students with SEN-status, which could partly explain the differential effect, for instance, trough available teacher resources or peer spillover effect (see also Hienonen et al., 2018). Since the class context seems to be more important for the low achievers, more classroom resources should be allocated particularly to these classes, where the overall effect of schooling seems to be bigger.

Although the differences between Finnish schools have been one of the lowest in the OECD countries (OECD, 2001; 2007; 2008; 2016), there seems to be considerable variation between classrooms (see also Berisha & Seppänen, 2017; Hotulainen, 2016; Yang Hansen et al., 2014). Besides classroom and peer effects, differences between classes may be the result of the initial class placement practices, through which evenly-achieving students are placed together more often. These between-class differences in achievement seem to increase as students progress through school years, leading to even more unequal settings and to concern especially the lower-achieving students and classrooms. Although classroom patterns with diverging intraclass correlations provided important information, future research should explore the causal pathways and the differential ways that individual- and class-level factors impact student outcomes more profoundly. However, as was indicated in the present study, the fact that divergent classroom patterns exist may provide a starting point for such investigation.

4.3. Conclusions

The application of a classroom perspective to the current study and its focus on the contextual effects provides a picture of classrooms as complex learning environments in which students are motivated and perform as a function of their own beliefs and ability and their participation in broader groups of classmates. By using MLPA we were able to reveal the impact of contextual effects on groups of students, an impact that is not usually acknowledged in pattern-centred (or person-oriented) approach. In light of our findings, it is suggested that models profiling students based on achievement be extended to include factors beyond the individual, such as the role of classroom membership or other contextual factors. Understanding the distinct classroom patterns may have critical implications for preventing declining performance or undesirable test taking behaviour, as different classes may require different types of support to maintain students' engagement in school and keep their achievement high. Our study and the analytical method chosen focuses attention on the importance of examining both student and classroom patterns and highlights the need for more research that is focused on classroom differences in the development of achievement during adolescence. The findings of the present study suggest that the classroom membership may create class level quality differences in both the preconditions and the development of learning.

Declaration of Competing Interest

None.

Acknowledgements

This study was supported by the Academy of Finland project "Redefining adolescent learning: A multi-level longitudinal cohort study of adolescent learning, health, and well-being in educational transitions in Finland" (288208). Elina E. Ketonen would like to thank the Faculty of Educational Sciences, University of Helsinki, for her mobility funding and to express her gratitude to the Department of Education at the University of Oxford for hosting her during her Fellowship at Oxford.

References

Acacio-Claro, P. J., Doku, D. T., Koivusilta, L. K., & Rimpelä, A. H. (2018). How socio-economic circumstances, school achievement and reserve capacity in adolescence predict adult education level: A three-generation study in Finland. *International Journal of Adolescence and Youth, 23*(3), 382–397.

Altermatt, E. R., & Pomerantz, E. M. (2003). The development of competence-related and motivational beliefs: An investigation of similarity and influence among friends. *Journal of Educational Psychology, 95*(1), 111–123.

Asparouhov, T., & Muthen, B. (2008). Multilevel mixture models. In G. R. Hancock, & K. M. Samuelsen (Eds.). *Advances in latent variable mixture models* (pp. 27–51). Charlotte, NC: Information Age.

Attali, Y. (2016). Effort in low-stakes assessments: What does it take to perform as well as in a high-stakes setting?. *Educational and Psychological Measurement, 76*(6), 1045–1058.

Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology, 43*(1), 272–311.

Bergman, L. R., & Andersson, H. (2010). The person and the variable in developmental psychology. *Zeitschrift für Psychologie, 218*, 155–165.

Berisha, A. K., & Seppänen, P. (2017). Pupil selection segments urban comprehensive schooling in Finland: Composition of school classes in pupils' school performance, gender, and ethnicity. *Scandinavian Journal of Educational Research, 61*(2), 240–254.

Bowers, A. J., & Sprott, R. (2012). Examining the multiple trajectories associated with dropping out of high school: A growth mixture model analysis. *The Journal of Educational Research, 105*(3), 176–195.

Brophy, J. E. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology, 75*(5), 631–661.

Cole, J. S., Bergin, D. A., & Whittaker, T. A. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology, 33*, 609–624.

Collins, C., & Gan, L. (2013). *Does sorting students improve scores?* Retrieved fromNational Bureau of Economic Research: An analysis of class composition.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analysis Archives, 8*(1), 1–44.

De Fraine, B., Van Damme, J., Van Landeghem, G., Opdenakker, M. C., & Onghena, P. (2003). The effect of schools and classes on language achievement. *British Educational Research Journal, 29*(6), 841–859.

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: Unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology, 110*(8), 1112–1126.

Eurydice. (2018). The Organisation of School Time in Europe. Primary and General Secondary Education – 2018/19. Eurydice Facts and Figures. Luxembourg: Publications Office of the European Union.

Fagginger Auer, M. F., Hickendorff, M., Van Putten, C. M., Béguin, A. A., & Heiser, W. J. (2016). Multilevel latent class analysis for large-scale educational assessment data: Exploring the relation between the curriculum and students' mathematical strategies. *Applied Measurement in Education, 29*(2), 144–159.

Finnish National Board of Education (2004). *National Core Curriculum for Basic Education 2004.* Helsinki: Finnish National Board of Education.

Finnish National Board of Education (2014). *National Core Curriculum for Basic Education 2014.* Helsinki: Finnish National Board of Education.

Flunger, B., Trautwein, U., Nagengast, B., Lüdtke, O., Niggli, A., & Schnyder, I. (2015). The Janus-faced nature of time spent on homework: Using latent profile analyses to predict academic achievement over a school year. *Learning and Instruction, 39*, 97–106.

Forgasz, H. (2010). Streaming for mathematics in Victorian secondary schools. *Australian Mathematics Teacher, 66*(1), 31–40.

Geiser, C., Lehmann, W., & Eid, M. (2010). Separating" rotators" from" nonrotators" in the mental rotations test: A multigroup latent class analysis. *Multivariate Behavioral Research, 41*(3), 261–293.

Gibb, S. J., Fergusson, D. M., & Horwood, J. L. (2008). Gender differences in educational achievement to age 25. *Australian Journal of Education, 52*(1), 63–80.

Glock, S., & Krolak-Schwerdt, S. (2013). Does nationality matter? The impact of stereotypical expectations of student teachers' judgments. *Social Psychology of Education, 16*(1), 111–127.

Goldenberg, C. (1992). The limits of expectations: A case for case knowledge about teacher expectancy effects. *American Educational Research Journal, 29*(3), 517–544.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2004). Why public schools lose teachers. *Journal of Human Resources, 39*(2), 326–354.

Harker, R., & Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement, 15*(2), 177–199.

Hart, S. A., Logan, J. A. R., Thomson, L., Kovas, Y., McLoughlin, G., & Petrill, S. A. (2016). A latent profile analysis of math achievement, numerosity, and math anxiety in twins. *Journal of Educational Psychology, 108*(2), 181–193.

Harter, S. (2012). *The Construction of the Self: Developmental and Sociocultural Foundations* ((2nd ed.).). New York & London: The Guilford Press.

Hautamäki, J., Kupiainen, S., Marjanen, J., Vainikainen, M.-P., & Hotulainen, R. (2013). Oppimaan oppiminen peruskoulun päättövaiheessa: Tilanne vuonna 2012 ja muutos vuodesta 2001 [Learning to learn at the end of basic education - results in 2012 and changes from 2001]. Department of Teacher of Education Research Report No 347. Helsinki: University of Helsinki.

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*(1), 60–87.

Henry, K. L., & Muthén, B. (2010). Multilevel latent class analysis: An application of adolescent smoking typologies with individual and contextual predictors. *Structural Equation Modeling, 17*(2), 193–215.

Hickendorff, M., Heiser, W. J., Van Putten, C. M., & Verhelst, N. D. (2009). Solution strategies and achievement in Dutch complex arithmetic: Latent variable modeling of change. *Psychometrika, 74*(2), 331–350.

Hienonen, N., Lintuvuori, M., Jahnukainen, M., Hotulainen, R., & Vainikainen, M. P. (2018). The effect of class composition on cross-curricular competences–Students with special educational needs in regular classes in lower secondary education. *Learning and Instruction, 58*, 80–87.

Hotulainen, R. (2016). Osaamisen kehittyminen kuntatyypeittäin, kouluittain, luokittain ja osaamisprofiileittain [Development of achievement by municipalities, schools, classrooom and competence profiles]. In R. Hotulainen (Ed.). *Learning and well-being: Transition from lower to upper secondary education in the Helsinki Metropolian Area* (pp. 97–110). Helsinki: Helsingin yliopisto.

Jahnukainen, M., & Itkonen, T. (2015). Tiered intervention: History and trends in Finland and the United States. *European Journal of Special Needs Education, 31*(1), 140–150.

Jerrim, J. (2016). PISA 2012: How do results for the paper and computer tests compare? *Assessment in Education: Principles, Policy & Practice, 23*(4), 495–518.

Jussim, L., & Harber, K. D. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review, 9*(2), 131–155.

Ketonen, E., Haarala-Muhonen, A., Hirsto, L., Hänninen, J. J., Wähälä, K., & Lonka, K. (2016). Am I in the right place? Academic engagement and study success during the first years at university. *Learning and Individual Differences, 51*, 141–148.

Kindermann, T. A. (2007). Effects of naturally existing peer groups on changes in

academic engagement in a cohort of sixth graders. *Child Development, 78*(4), 1186–1203.
Koivuhovi, S., Vainikainen, M. P., Kalalahti, M., & Niemivirta, M. (2019). Changes in children's agency beliefs and control expectancy in classes with and without a special emphasis in finland from grade four to grade six. *Scandinavian Journal of Educational Research, 63*(3), 427–442.
Kosunen, S., Bernelius, V., Seppänen, P., & Porkka, M. (2016). School choice to lower secondary schools and mechanisms of segregation in Urban Finland. *Urban Education,* 1–28.
Lam, C. (2014). Estimating Various Kinds of Peer Effects on Academic Performance (Working Paper). Department of Economics, University of Chicago. Available: http://www.albany.edu/economics/research/seminar/files/chungsang.pdf.
Lukočienė, O., & Vermunt, J. K. (2010). Determining the number of components in mixture models for hierarchical data. In A. Fink, L. Berthold, W. Seidel, & A. Ultsch (Eds.). *Advances in data analysis, data handling and business intelligence* (pp. 241–249). Berlin, Heidelberg, Germany: Springer.
Makara, K. A., & Madjar, N. (2015). The role of goal structures and peer climate in trajectories of social achievement goals during high school. *Developmental Psychology, 51*(4), 473–488.
Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J., Abduljabbar, A. S., et al. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist, 47*(2), 106–124.
Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., et al. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research, 44*(6), 764–802.
McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context and the achievement gap. *Journal of School Psychology, 46*(3), 235–261.
Muthén, L. K., & Muthén, B. O. (1998–2018). Mplus user's guide (8th ed.). Los Angeles, CA: Muthén & Muthén.
Mäkikangas, A., Tolvanen, A., Aunola, K., Feldt, T., Mauno, S., & Kinnunen, U. (2018). Multilevel latent profile analysis with covariates: Identifying job characteristics profiles in hierarchical data as an example. *Organizational Research Methods, 21*(4), 931–954.
Nash, R. (2003). Is the school composition effect real? A discussion with evidence from the UK PISA data. *School Effectiveness and School Improvement, 14*(4), 441–457.
OECD (2001). *Knowledge and skills for life: First results from the OECD programme for international student assessment (PISA) 2000.* Paris: OECD.
OECD (2007). *Science competencies for tomorrow's world: Results from PISA 2006.* Paris: OECD.
OECD (2008). *Education at a glance 2008: OECD indicators.* Paris: OECD.
OECD. (2013). PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices (Volume IV). PISA, OECD Publishing. http://doi.org/10.1787/9789264201156-en.
OECD (2016). PISA 2015 Results (Volume I): Excellence and Equity in Education. *PISA, OECD Publishing.*. https://doi.org/10.1787/9789264266490-en.
Official Statistics of Finland (OSF). (2017). Special education [e-publication]. Helsinki: Statistics Finland. Retrieved June 23, 2017, from http://www.stat.fi/til/erop/index_en.html.
Peetsma, T., van der Veen, I., Koopman, P., & van Schooten, E. (2006). Class composition influences on pupils' cognitive development. *School Effectiveness and School Improvement, 17*(3), 275–302.
Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40.
Prinstein, M. J., & Dodge, K. A. (2008). *Understanding peer influence in children and adolescents.* Guilford Press.
Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (Vol. 1). Sage.
Reichelt, M., Collischon, M., & Eberl, A. (2019). School tracking and its role in social reproduction: Reinforcing educational inheritance and the direct effects of social origin. *British Journal of Sociology*. https://doi.org/10.1111/1468-4446.12655.
Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectations and pupils' intellectual development.* New York, NY: Holt, Rinehart and Winston.
Rubie-Davies, C. (2009). Teacher Expectations and Labeling. In L. J. Saha, & A. G. Dworkin (Vol. Eds.), *International Handbook of Research on Teachers and Teaching. Springer International Handbooks of Education: vol. 21*, (pp. 695–707). Boston, MA: Springer.
Rubie-Davies, C. M. (2010). Teacher expectations and perceptions of student attributes: Is there a relationship? *British Journal of Educational Psychology, 80*, 121–135.
Ryan, A. M. (2001). The peer group as a context for the development of young adolescent motivation and achievement. *Child Development, 72*(4), 1135–1150.
Sammons, P., Nuttall, D., & Cuttance, P. (1993). Differential school effectiveness: Results from a reanalysis of the Inner London Education Authority's Junior School Project data. *British Educational Research Journal, 19*(4), 381–405.
Shin, H., & Ryan, A. M. (2014). Early adolescent friendships and academic adjustment: Examining selection and influence processes with longitudinal social network analysis. *Developmental Psychology, 50*(11), 2462–2472.
Slavin, R. (1990). Achievement effects of ability grouping in secondary schools: A best evidence synthesis. *Review of Educational Research, 60*(3), 471–499.
Stäbler, F., Dumont, H., Becker, M., & Baumert, J. (2017). What happens to the fish's achievement in a little pond? A simultaneous analysis of class-average achievement effects on achievement and academic self-concept. *Journal of Educational Psychology, 109*(2), 191–207.
Thapa, A., Cohen, J., Guffey, S., & Higgins-D'Alessandro, A. (2013). A review of school climate research. *Review of Educational Research, 83*(3), 357–385.
Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research, 37*(5), 483–504.
Thuneberg, T., Hautamäki, J., & Hotulainen, R. (2015). Scientific reasoning, school achievement and gender: A multilevel study of between and within school effects in Finland. *Scandinavian Journal of Educational Research, 59*(3), 337–356.
Thuneberg, H., Vainikainen, M. P., Ahtiainen, R., Lintuvuori, M., Salo, K., & Hautamäki, J. (2013). Education is special for all: The Finnish support model. *Gemeinsam Leben, 21*(2), 67–78.
Tieso, C. (2003). Ability grouping is not just tracking any more. *Roeper Review, 26*(1), 29–37.
Timmermans, A. C., Kuyper, H., & van der Werf, G. (2015). Accurate, inaccurate, or biased teacher expectations: Do Dutch teachers differ in their expectations at the end of primary education? *British Journal of Educational Psychology, 85*(4), 459–478.
Timmermans, A. C., & Rubie-Davies, C. M. (2018). Do teachers differ in the level of expectations or in the extent to which they differentiate in expectations? Relations between teacher-level expectations, teacher background and beliefs, and subsequent student performance. *Educational Research and Evaluation, 24*(3–5), 241–263.
Van Barneveld, C., Pharand, S.-L., Ruberto, L., & Haggarty, D. (2013). Student motivation in large-scale assessments. In M. Simon, K. Ercikan, & M. Rousseu (Eds.). *Improving Large-Scale Assessment in Education. Theory, Issues and Practice* (pp. 43–61). Thousand Oaks, CA: Sage Publications.
Van den Broeck, A., Vansteenkiste, M., De Witte, H., Soenens, B., & Lens, W. (2010). Capturing autonomy, competence, and relatedness at work: Construction and initial validation of the Work-Related Basic Need Satisfaction Scale. *Journal of Occupational and Organizational Psychology, 83*(4), 981–1002.
Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved threestep approaches. *Political Analysis, 18*, 450–469.
Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. Hagenaars, & A. McCutcheon (Eds.). *Applied latent class analysis* (pp. 89–106). Cambridge, UK: Cambridge University Press.
Vermunt, J. K., & Magidson, J. (2013). *Latent GOLD 5.0 upgrade manual.* Belmont, MA: Statistical Innovations.
Warburton, V. E. (2017). Peer and teacher influences on the motivational climate in physical education: A longitudinal perspective on achievement goal adoption. *Contemporary Educational Psychology, 51*, 303–314.
Wentzel, K. R., Muenks, K., McNeish, D., & Russell, S. (2017). Peer and teacher supports in relation to motivation and effort: A multi-level study. *Contemporary Educational Psychology, 49*, 32–45.
Willms, J. D. (1985). The balance thesis: Contextual effects of ability on pupils' O-grade examination results. *Oxford Review of Education, 11*(1), 33–41.
Yang Hansen, K., Gustafsson, J.-E., & Rosén, M. (2014). School performance difference and policy variations in Finland, Norway and Sweden. In K. Yang Hansen (Ed.). *Northern Lights on TIMMS and PIRLS 2011* (pp. 25–48). Copenhagen: Nordic Council of Ministers.